

Kurzinformationen zu den präsentierten Programmsystemen

CYRIL BELICA

COSMAS I im Internet

Dies ist eine Beta-Version einer WWW-Schnittstelle zum COSMA-I Recherchesystem. Zur Zeit ist nur eine sehr eingeschränkte Auswahl der COSMAS-I-Funktionen über diese Schnittstelle verfügbar. Eine neue COSMAS-Version wird implementiert und sollte bald freigegeben werden.

Die über COSMAS I recherchierbaren Korpora umfassen zur Zeit ca. 484 Millionen Textwörter. Die Auswahl der öffentlich verfügbaren Korpora ist aus urheberrechtlichen Gründen auf ca. 128 Millionen Textwörter beschränkt. Die Dauer einer anonymen Recherche-Sitzung ist auf 60 Millionen limitiert. Beachten Sie bitte, dass COSMAS I keine „Internet-Suchmaschine“, sondern ein Korpus-Recherche- und -Analyse-Werkzeug ist und daher über eine eigene, auf korpuslinguistische Bedürfnisse abgestimmte Suchabfragesprache verfügt.

Die kommerzielle Nutzung der Rechercheergebnisse ist nicht erlaubt.
<http://www/kt/cosmas.html>

FRANCK BODMER

Das Recherchieren in Gesprächstranskripten mit COSMAS II

COSMAS II ist die Nachfolgesoftware von COSMAS I für die Verwaltung und das Recherchieren in den annotierten Korpora des IDS. Es erlaubt außerdem das Recherchieren in Transkripten, die aus Audio-Aufnahmen gewonnen wurden und mit der Aufnahme gekoppelt sind, so dass man sich anhand der gefundenen Treffer die Originalpassagen anhören kann.

Im Gegensatz zu den Korpora geschriebener Sprache ist bei Transkripten die Linearität des Textes nicht gegeben: Der Text ist auf Äußerungen (Turns) mehrerer Sprecher verteilt, die gleichzeitig sprechen (Simultanpassagen) und sich gegenseitig unterbrechen (Sprecherwechsel, Einschü-

be) können. Durch diese Phänomene werden Wörter außerdem fragmentiert (Wortfragmente).

Diese Besonderheiten gesprochener Sprache werden in COSMAS II berücksichtigt, verwaltet und recherchierbar gemacht. Z. B. ist ungeachtet von Einschüben eine Wortabstandssuche über einen Sprecher möglich, eine Wortsuche trotz potentieller Fragmentierung durchführbar und das Abfragen von gleichzeitig Gesprochenem formulierbar. Dazu kommen eine Reihe weiterer recherchierbarer Annotationen wie Pausen, vokalisierte, in den Transkripten nicht lexikalisierte Laute, Transkriptionskommentare und unverständliche oder undeutliche Stellen.

COSMAS II befindet sich derzeit in einem prototypischen Zustand, in welchem die Transkripte noch als SGML-Dokumente angezeigt werden. Zur besseren Visualisierung soll COSMAS II in Zukunft mit einem spezialisierten Transkript-Editor gekoppelt werden. So ist es z. B. grundsätzlich möglich, Transkripte, die mit dem spezialisierten DIDA-Editor erzeugt wurden, nach einer Recherche in COSMAS II mit Hilfe von DIDA in der Partiturdarstellung zu präsentieren.

<http://www.ids-mannheim.de/zdv/cosmas2>

MELANIE BRINKSCHULTE

Der computerbasierte C-Test als Einstufungstest

Seit dem WS 94/95 qualifizieren sich Studierende am Sprachenzentrum der Westfälischen Wilhelms-Universität mittels eines zentralisierten Einstufungstests für studienbegleitende Sprachkurse. Der für die Kursanmeldung obligatorische Test erfolgt computerbasiert.

Bei dem von Klein-Braley/Raatz (1982) entwickelten Cloze-Test handelt es sich um einen Proficiencytest, der lediglich ein globales Maß an Sprachkompetenz misst und nicht einzelne Stärken oder Schwächen des Probanden herausfinden kann. Der Vorteil dieses Testverfahrens liegt in seiner Effizienz, da der Test relativ unaufwendig zu entwickeln ist und wenig Zeit für die Administrierung und Auswertung beansprucht.

Am Sprachenzentrum der WWU können Studierende zu Beginn eines jeden Semesters den C-Test in den Sprachen Deutsch, Englisch, Französisch, Italienisch, Russisch und Spanisch absolvieren. Für jede Sprache wird aus einem Pool eine Zufallskombination von vier kleineren Texten ausgewählt, die jeweils 25 Lücken aufweisen. Entsprechend der Systematik der Cloze-Tests fehlt bei jedem zweiten Wort die zweite Hälfte, die von dem Prüfling ausgefüllt werden muss. Auf diese Weise wird das globale

Sprachverstehen gemessen. Untersuchungen haben ergeben, dass Muttersprachler über die Kompetenz verfügen, alle Lücken angemessen schließen zu können (Grotjahn 1995). Den Probanden steht eine Zeit von 25 Minuten zur Verfügung, um den Test zu absolvieren. Hiernach erhalten sie eine automatische Auswertung mit ihrer erreichten Punktzahl, die sie für die Anmeldung zu verschiedenen Kursstufen benötigen.

Durch die Anwendung des computerbasierten Cloze-Tests konnten im letzten Semester (WS98/99) über 3000 Kursanmeldungen berücksichtigt werden, so dass die KursinteressentInnen ihrem jeweiligen Kursniveau zugeordnet werden konnten.

Literaturhinweise:

- Grießhaber, W. (1999): Der C-Test als Einstufungstest. Münster (mimeo)
Grotjahn, R. (1995): Der C-Test. State of the Art. In: Zeitschrift für Fremdsprachenforschung 6 (2)/1995, S. 37–60.
Klein-Braley, C./Raatz, U. (1982): Der C-Test: ein neuer Ansatz zur Messung allgemeiner Sprachbeherrschung. In: AKS-Rundbrief Heft 4/1982, S. 23–37.

THOMAS BURCH/JOHANNES FOURNIER/KURT GÄRTNER

Mittelhochdeutsche Wörterbücher auf CD-ROM und im Internet

Die derzeit vorhandenen, wichtigsten Wörterbücher zum Mittelhochdeutschen sind eng aufeinander bezogen, ergänzen sich gegenseitig und müssen stets gemeinsam benutzt werden. Eine an der Universität Trier im Rahmen eines DFG-Projektes erstellte elektronische Fassung dieser Wörterbücher erleichtert ihre gemeinsame Benutzung dadurch, dass die komplexen Verweisstrukturen auf Hyperlinks abgebildet worden sind. Darüber hinaus ermöglicht eine metasprachliche, SGML/TEI-konforme Auszeichnung lexikographischer Information den Aufbau einer Wörterbuchdatenbank, mit deren Hilfe systematische Abfragen an den Inhalt der mittelhochdeutschen Wörterbücher möglich werden.

Wesentlich für die Konzeption des digitalen Wörterbuchverbundes ist ferner die Einbindung von PostScript-Files der Wörterbuchseiten, die den Bezug zum Buch auch vom Bildschirm aus jederzeit garantieren, sowie die Einbeziehung eines digitalen Quellenverzeichnisses, das nicht nur die Auflösung der Siglen ermöglicht und zu den als Quellen benutzten Textausgaben hinführt, sondern auch Informationen zur Datierung, zur Schreibsprache und zu den Textsorten der Quellen enthält, die für syste-

matische Abfragen an das elektronische Wörterbuch herangezogen werden können.

<http://gaer27.uni-trier.de/MWV-online/MWV-online.html>.

GREGOR ERBACH

Demonstrationen der Systeme MULINEX, TWENTYONE und DIET

Systemvorführungen des Forschungsbereich Sprachtechnologie des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI, Universität Saarbrücken)

Informationen über den Forschungsbereich: <http://www.dfki.de/lt/>

Im Projekt MULINEX haben mehrere europäische Firmen unter Koordination des DFKI ein System entwickelt, das die multilinguale Suche und Navigation auf dem WWW und in Intranet ermöglicht. Die Sprache der Suchschnittstelle kann vom Benutzer eingestellt werden. Suchanfragen werden übersetzt. Mit dem „Anfrageassistent“, lassen sich die Übersetzungen überprüfen und anpassen. Suchergebnisse werden mit Sprache, Themenkategorien und einer Zusammenfassung gezeigt. Die Ergebnisse können automatisch nach Sprache und Kategorie gefiltert werden. Die gefundenen Webseiten können durch ein integriertes kommerzielles Übersetzungssystem übersetzt werden.

<http://www.dfki.de/pas/f2w.cgi?ltp/mulinex-g>

Im Projekt TWENTYONE hat das DFKI mit Umweltschutzorganisationen, Technologiefirmen und Forschungsinstituten aus verschiedenen europäischen Ländern zusammengearbeitet, um Dokumente aus dem Umweltschutzbereich, insbesondere zum Thema „nachhaltige Entwicklung“ auf CD-ROM und im Internet verfügbar zu machen.

Dokumente, die nicht elektronisch sondern in Papierform vorliegen, wurden eingescannt und umgewandelt. Die Dokumentstruktur wurde automatisch analysiert, und der Textinhalt mittels optischer Zeichenerkennung (OCR) extrahiert. Durch Fuzzy Matching bei der Suche ist das System robust gegenüber Schreibfehlern in den Dokumenten und in der Anfrage sowie gegenüber OCR-Fehlern. Durch linguistische Analyse der Texte kann nicht nur nach Einzelwörtern gesucht werden, sondern auch nach wichtigen Phrasen. Die Texte wurden vor dem Indizieren automatisch übersetzt,

so dass in drei Sprachen (Deutsch, Englisch, Französisch) gesucht werden kann.

<http://www.dfki.de/pas/f2w.cgi?ltp/twentyone-g>

In dem vom DFKI koordinierten EU-Projekt DIET haben europäische Firmen und Forschungszentren eine umfassende Umgebung für die Konstruktion, Annotation und Pflege von strukturierten Referenzdaten für die Diagnose und Evaluierung von sprachtechnologischen Anwendungen entwickelt. Das System ist in einer konfigurierbaren, offenen Client/Server-Architektur entwickelt. Der Benutzer kann Daten konstruieren und manuell oder automatisch annotieren, indem er aus einer Menge von gegebenen Annotations-Datentypen auswählt, die mit allen notwendigen Funktionen assoziiert sind. Die umfangreichen strukturierten Testdaten repräsentieren linguistische Phänomene auf den Ebenen Morphologie, Syntax und Diskurs. Die Annotationen decken linguistische und anwendungsspezifische Aspekte ab, und unterstützen dadurch sowohl die Transparenz der Daten als auch den optimalen Zugriff. Für die Anpassung der Daten an neue Domänen werden dem Benutzer verschiedene Customisierungsmöglichkeiten geboten. So können durch einen Textprofiling-Prozess die kontrollierten Daten einer Testsuite mit entsprechenden Konstruktionen in domänenspezifischen Korpora in Verbindung gebracht werden. Mithilfe von lexikalischen Ersetzungsfunktionen kann der Benutzer das Vokabular der Testitems an domänenspezifische Terminologie anpassen. Mithilfe der Datenbank kann der Benutzer Ergebnisse verschiedener Test- und Evaluierungszyklen speichern und verwalten.

<http://www.dfki.de/pas/f2w.cgi?ltp/diet-g>

MARCUS FACH/MARCEL SCHILLING/RUDOLF SCHMIDT/
WILFRIED SCHÜTTE

Maschinelle Segmentierung von Korpora gesprochener Sprache des IDS (Alignment-Projekt)

Ziel dieses Projektes ist ein automatisches Alignment von Daten gesprochener Sprache des Instituts für Deutsche Sprache (IDS) sowie die Weiterentwicklung der am IDS und am Lehrstuhl für Experimentelle Phonetik, Institut für Maschinelle Sprachverarbeitung (IMS) an der Universität Stuttgart dazu vorhandenen Werkzeuge (Aligner).

Unter Alignment ist die Synchronisation von Sprachsignal und zugehörigem Text zu verstehen. Die akustischen Schwingungen des Sprachsignals werden zeitsynchron den entsprechenden Wörtern einer zuvor erstellten Transliteration zugeordnet. Für das IDS bietet diese Repräsentation die einfachste und schnellste Möglichkeit, die direkte Abbildung von Text auf Sprachsignale in die dort entwickelte Datenbank COSMAS II zu integrieren. Die Bereitstellung so immenser maschinenlesbarer Daten stellt eine wesentliche Voraussetzung für die Weiterentwicklung der Sprachtechnologien wie Sprachsynthese und Spracherkennung am IMS dar.

Beschreibung des Projekts:

Das IDS verfügt über große Mengen an gesprochenem Deutsch aus den verschiedensten dialektalen Bereichen und Aufnahmesituationen (Aufnahmen aus Fernsehen, Radio und „Feldforschung“). Diese Aufnahmen liegen in unterschiedlichster Form und in unterschiedlicher Audioqualität vor. Ein großer Teil der Aufnahmen ist transkribiert, und die Menge der maschinenlesbaren Transkriptionstexte wächst fortlaufend.

Durch die Erstellung dieser maschinenlesbaren Transkriptionstexte ist es nun möglich, mittels des Aligners die Abbildung (Annotation) von Transkription auf die Sprachsignale automatisch durchzuführen. Dabei handelt es sich um eine Abbildung auf Wortebene, die es ermöglicht, mit einfachen Methoden der Textrecherche Ausschnitte aus einem beliebigen Sprachsignal zu erhalten. Ziel des weiteren Korpusaufbaus ist die Aufnahme von Text, digitalisiertem Ton und den Annotationen in die am IDS installierte Datenbank COSMAS II.

Aus Sicht des IDS kommt dem Alignment von natürlichsprachigen Gesprächskorpora eine strategische Bedeutung zu. Damit werden in neuartiger Weise automatische Recherchen und Auswertungen des Sprachkorpus möglich, die einem großen Spektrum von Untersuchungsinteressen dienen können. So können umfangreiche sprachwissenschaftliche Studien über die interne Struktur der deutschen Sprache einschließlich der Dialekte durchgeführt werden, für die bislang kein ausreichendes Sprachmaterial maschinell zugreifbar war.

Für das IMS stellt die Menge der am IDS automatisch alignierten und manuell überprüften Daten einen Fundus zum Training des dortigen Spracherkenners und Aligners dar. In diesem Zusammenhang werden am IMS neue Modelle (HMM) für sprachliche Phänomene entwickelt, die bei der Entwicklung des Aligners nicht vorhanden waren und folglich bei der Verarbeitung der Daten des IDS Schwierigkeiten bereiten. Der so verbesserte Aligner wird im IDS installiert und dort weiter betreut und angewendet.

<http://www.ids-mannheim.de/prag/alignment.html>

JÜRGEN HANDKE/FRAUKE INTEMANN

Die CD-ROM „Die interaktive Einführung in die Linguistik“

Die CD-ROM „Die interaktive Einführung in die Linguistik“ wurde am Fachbereich Neuere Fremdsprachen und Literaturen der Philipps-Universität Marburg entwickelt.

Dieses neuartige CBT-System stellt den ersten umfassenden Versuch dar, alle Inhalte einer Einführung in die Sprachwissenschaft multimedial aufzubereiten und eine traditionelle Lehrveranstaltung – inklusive der Leistungskontrolle – durch ein Selbstlernsystem zu ersetzen.

Die primäre Zielgruppe der CD-ROM sind Studierende der Germanistik im Grundstudium, die an der Veranstaltung „Einführung in die Linguistik“ teilnehmen. Studierenden anderer sprachwissenschaftlicher Fachrichtungen gibt die CD-ROM die Möglichkeit, ihr Wissen zu erweitern und dabei die Vorteile einer multimedialen Darstellung zu nutzen. Lehrenden bietet das System eine Fülle von Material, da linguistische Phänomene aufgrund ihrer Dynamik mit herkömmlichen Medien nicht adäquat dargestellt werden können.

Die Vorführung zeigt alle zentralen Inhalte der CD-ROM: Phonetik/Phonologie, Morphologie, Syntax und Semantik. Das System verfügt über einige zusätzliche Module, die nicht zu den Kerngebieten der Linguistik zählen. Das Modul ‚Wozu Linguistik?‘ soll den Studierenden verschiedene Motive für die Beschäftigung mit der Theorie der Sprache nahebringen. Unter dem Kapitel ‚Sprachen der Welt‘ finden sich drei Unterkapitel zu Sprachen und Sprachfamilien, zur Evolution des Menschen unter besonderer Berücksichtigung der Entwicklung der Sprachfähigkeit und zur Geschichte der deutschen Sprache.

Zu allen Bereichen gibt es allgemeine bibliographische Angaben, aber auch spezielle Hinweise auf Kapitel aus ausgewählten Büchern zum jeweiligen Thema.

Die Besonderheit des Systems ist der Interaktive Tutor. Zu allen Kerngebieten gibt es eine Fülle von Übungen, mit denen das erlernte Wissen trainiert und der Lernfortschritt überprüft werden kann. Neben Multiple-Choice-Übungen gibt es eine Reihe anderer Übungsformen wie Erkennungsaufgaben (Phoneme, Morpheme, Phrasen, ...), Konstruktionsaufgaben (Syntax-Bäume), Analysen, Benennungsaufgaben und vieles mehr. Der Lernfortschritt der Studierenden wird dabei individuell protokolliert, einige Übungen bieten mehrere, dem Lernfortschritt angepasste Schwierigkeitsstufen an.

Die CD-ROM ist im Max Hueber Verlag, München erschienen. Preis: 78,- DM

ULRIKE HASS-ZUMKEHR/CLAUDIA FRAAS/
CYRIL BELICA/SONJA MÜLLER-LANDMANN/
KATHRIN STEYER

LEKSIS – wissen über wörter

LEKSIS ist ein neues Projekt des Instituts für Deutsche Sprache und wird als computergestütztes Informationssystem zum deutschen Wortschatz konzipiert. Es baut nicht auf irgendeinem vorhandenen Wörterbuch auf, sondern soll Wissen zu Lexik, Lexikologie und linguistischen Termini als Hypertext organisieren und so flexibel und selektiv abrufbar machen. LEKSIS wird für lexikologische Untersuchungen unterschiedlichster Art eine umfassende und solide Datengrundlage bieten und über Internet für LinguistInnen und Nicht-LinguistInnen zugänglich sein. Der Aufbau von LEKSIS führt in innovative Bereiche der linguistischen Forschung, denn er berührt Probleme der Vernetzung von Wortschatzwissen, der Klassifikation von Wortschatz, der Repräsentation von lexikalischem Wissen vor dem Hintergrund übergeordneter komplexer Wissenszusammenhänge und der Strukturierung und Markierung von Information.

Siehe auch: Deutsche Sprache 1/99: „Vom Wörterbuch zum lexikalischen Informationssystem“

<http://www.ids-mannheim.de/leksis/>

HANS JÜRGEN HERINGER

Die multimedialen Trainingsprogramme „Verstehen. Eine Selbstanleitung“ und „Aus Fehlern lernen“

Verstehen

Es handelt sich um ein multimediales Trainingsprogramm, das intrakulturelles und interkulturelles Verstehen zum Thema hat. Zentrales Modul ist ein personen-zentriertes Lernprogramm, das Methoden vorführt und erprobt, mit denen man seine Verstehensfähigkeiten verbessern kann. Das Programm ist experienziell orientiert.

Aus Fehlern lernen

Dieses multimediale Lern- und Übungsprogramm basiert auf 8000 Lerndefehlern, die überwiegend in einem Projekt des Goethe-Instituts erho-

ben wurden. Ein Modul ist für Lehrer gedacht und erschließt systematisch die Fehlerdaten. Zusätzlich gibt es ein Modul für Lerner, das auf einem repräsentativen Sample von 600 Fehlern aufbaut. Alle Fehler sind mit Korrekturvorschlägen und einer Diagnose versehen.

Eine Referenzgrammatik und eine kontextsensitive Lernergrammatik vervollständigen das Programm.

http://www.phil.uni-augsburg.de/phil2/faecher/germanis/DaF_Forschung.htm

MONIKA KOLVENBACH/NORBERT VOLZ

AMADES – Arbeitspapiere und Materialien zur deutschen Sprache

AMADES – Arbeitspapiere und Materialien zur deutschen Sprache ist eine neue Reihe, die im Eigenverlag des IDS erscheint. Die Reihe will weg von der Veröffentlichungsform „nur Buch“ und da, wo es sich anbietet, hin zu multimedialen Veröffentlichungsformen, die den Umgang mit dem dargebotenen Inhalt erleichtern und den Arbeitsgewohnheiten der Benutzerinnen und Benutzer entgegenkommen. Wer für seine Arbeit „Papier zwischen zwei Deckeln“ bevorzugt, soll dies ebenso erhalten wie jemand, der sich lieber die Möglichkeiten elektronischer Medien nutzbar machen und z. B. über Hypertext schnell Verweise und Zusammenhänge realisieren will oder gar Bildaufnahmen – entweder als eigenständiges Video oder zusammen mit dem darauf Bezug nehmenden Text auf CD-ROM – zu veröffentlichen gedenkt.

Da AMADES sich als „deliver on demand“-Reihe versteht, ist es der Redaktion möglich, je nach Anfrage die gewünschte Form auszuliefern, da jeweils die entsprechenden – für die Veröffentlichung am besten geeigneten – Versionen zur Vervielfältigung vorliegen.

In dieser Reihe veröffentlichen hauptsächlich Mitarbeiterinnen und Mitarbeiter des IDS; sie ist aber auch – die Zustimmung der Herausgeber der Reihe (Leitungskonferenz des IDS) vorausgesetzt – für Arbeiten der o. a. Art von „Nicht-IDSlern“ offen.

Die Vorführung gibt einen Überblick über die bislang realisierten Projekte und die Möglichkeiten für zukünftige Veröffentlichungen in gedruckter und elektronischer Form.

<http://www.ids-mannheim.de/pub/amades>

SABINE KRÄMER-NEUBERT

Der Sprachatlas von Unterfranken multimedial

Am Beispiel der Frage „Wo wird das Geschirr aufbewahrt?“ wird ein neuer Weg aufgezeigt, der sowohl die Arbeit im Sprachatlas von Unterfranken (SUF) als auch konkrete Ergebnisse aus dem SUF im WWW präsentiert. Dies geschieht mit einer computeranimierten Projektion.

Nehmen Sie sich 5 Minuten Zeit und erleben Sie die Dialektologie in einem modernen Gewand! Auf einem separaten Rechner kann jeder Interessierte die SUF-Homepage besuchen. Dort finden sie Sprachkarten mit Tonbelegen, zu denen man sich per Mausklick weiterführende Informationen (Kommentar, Belegliste, Legende) anzeigen lassen kann.

Besuchen Sie uns im WWW und nehmen Sie Anteil am Fortgang unseres Sprachatlasprojekts!

<http://www.uni-wuerzburg.de/germanistik/spr/suf>

Asterix bei den Dialektologen

Unser Experiment, Dialekte anhand der Asterix-Mundart-Ausgaben zu untersuchen, wies diese Comics als hochkarätige sprachwissenschaftliche Forschungsquelle aus. Welche Stilmittel werden angewendet, um im Dialekt Sprachkomik zu erzeugen? Wie kreativ wird in eine Mundart übersetzt? Vom Fluchen und Schimpfen über Essen und Trinken bis hin zum Streiten und Kämpfen: Wie diese Tätigkeiten z. B. im Niederdeutschen realisiert werden, erfahren Sie unter

<http://members.tripod.de/dialektasterix/>

ULRICH SCHMITZ

LINSE – Ein Linguistik-Server im Internet

LINSE (<http://www.linse.ni-essen.de>) ist ein Internet-Server mit vielfältigem Angebot zu Sprache, Sprachwissenschaft, Sprachdidaktik, Neuen Medien und verwandten Gebieten. LINSE wendet sich an jede(n), die oder der sich für sprach- und medienbezogene Fragen interessiert, insbesondere aber an Sprachwissenschaftler, Linguistik-Studenten, Lehrer und Schüler, und erfüllt dabei mehrere verschiedene Aufgaben.

1. Sie bietet völligen Anfängern (vor allem aus dem Hochschulbereich) einen ersten Einstieg letzten Endes in das gesamte Internet.

2. Sie publiziert Aufsätze, kleine Schriften, Lernsoftware sowie Rezensionen von Büchern, CDs und Software.
3. Sie liefert Informationen, Bibliographien und Arbeitsmaterial für Forschung, Lehre und Unterricht.
4. Sie dient dem schnellen und direkten Austausch unter Wissenschaftlern, Studenten und allen Sprachinteressierten.
5. Sie ist ein Arbeitsmittel im Studium.

Literaturhinweis:

Cölfen, Elisabeth/Schmitz, Ulrich (1998): LINSE – ein Linguistik-Server im Internet. Angebot, Geschichte, akademische Ordnung und fröhliche Zukunft. In: Mitteilungen des Deutschen Germanistenverbandes 45, H. 3, S. 286–298

WOLFGANG SCHNEIDER/REINHARD FIEHLER/
PETER WAGENER

Datenbank „Gesprochenes Deutsch“

Computergestützte Erfassung und Erschließung der Tonaufnahmen des Deutschen Spracharchivs zum gesprochenen Deutsch

Das Projekt wird am IDS in der Abteilung Pragmatik durchgeführt. Es wird von der VW-Stiftung im Rahmen des Programms „Archive als Fundus der Forschung – Erfassung und Erschließung“ gefördert (Az.: II/72 759).

Das Projekt dient dazu, die 32 Korpora gesprochener Sprache, die im Deutschen Spracharchiv (DSAv) am IDS betreut werden (ca 15.000 Tonaufnahmen), besser zu dokumentieren und für Interessenten einen schnellen und einfachen Zugriff auf die Tonaufnahmen, Transkripte und Begleitmaterialien zu ermöglichen.

Die einzelnen Schritte der Projektarbeit sind:

- Einrichtung einer Informations-Datenbank, die jede einzelne Aufnahme durch einen definierten Satz von Informationen beschreibt.
- Einrichtung einer Volltext-Datenbank für die über 8.000 vorliegenden Transkripte zu den Tonaufnahmen.
- Synchronisierung der Audio-Files mit den entsprechenden Transkriptionen (Alignment).
- In synchronisierten Transkripten kann man von einer beliebigen Stelle im Text gezielt auf den zugehörigen Ausschnitt der Tonaufnahme zugreifen und ihn am Computer wiedergeben zu lassen.
- Suche nach Textbelegen in den Transkripten der Volltextdatenbank.

Dies wird durch das am IDS entwickelte Volltext-Retrieval-System COS-MAS II ermöglicht.

BRUNO STRECKER/ROMAN SCHNEIDER

GRAMMIS – Das grammatische Informationssystem des Instituts für Deutsche Sprache

GRAMMIS ist ein Projekt des Instituts für Deutsche Sprache in Mannheim, das zum Ziel hat, die ach so trockene, oft schwer verständliche deutsche Grammatik leichter zugänglich und soweit möglich, unterhaltsam und interessant zu machen. Zu diesem Zweck wird ein elektronisches Informationssystem aufgebaut, das – einmal fertiggestellt – diese Zugriffsmöglichkeiten bieten wird:

- Grammatisch wenig Vorbelastete, die gerade mal wissen wollen, wie dies oder jenes zu formulieren sein könnte, finden Hilfe bei einer Liste häufig gestellter Fragen zur Grammatik, und, wo das nicht genügt, ein grammatisches Wörterbuch und ein Suchleitsystem, das sie bei der genaueren Bestimmung ihres Problems unterstützt.
- Wer sich über den Einzelfall hinaus für knifflige grammatische Fragen interessiert, sollte gleich zu den Hauptschwierigkeiten der deutschen Grammatik gehen.
- Grammatisches Grundwissen in eher traditionell systematischer Form vermittelt die gleichnamige Einheit.
- Eine Grammatikexpertin bietet kompakte Erklärungen grammatischer Fachausdrücke und Übergänge zu detaillierteren Beschreibungen und einschlägiger Fachliteratur.
- Für Grammatikspezialisten und solche, die es werden wollen, bietet die Grammatik für Fortgeschrittene vertiefte Informationen zum Wie und Was unserer Beschreibungen und Erklärungen.
- Umfangreiche Bibliografien zur deutschen Grammatik, allgemein und zu speziellen Themenbereichen, sowie eine Sammlung maschinenlesbarer Texte zu grammatischen Themen runden das System ab.

Alle Komponenten von GRAMMIS werden über Hyperlinks verbunden, sodass man jederzeit Hilfe bei der Grammatikexpertin suchen kann, wenn man auf einen Fachausdruck stößt, der einem noch unbekannt ist, oder in der Bibliografie nachsehen kann, was an weiterführender Literatur zu dem Thema zu finden ist.

Damit man sich leicht und schnell orientieren kann, wird allen Informationseinheiten eine Kurzfassung vorangestellt, von der aus man mit einem Mausklick zu einer detaillierten Darstellung gelangt. Wer noch weiter einsteigen will, findet, wo dies angebracht erscheint, Übergänge zu vertieften Informationen.

Einige Komponenten von GRAMMIS können Sie jetzt schon nutzen, doch das meiste ist derzeit noch in Arbeit. Aber wir arbeiten daran. Schauen Sie von Zeit zu Zeit wieder bei uns vorbei. Sie werden immer wieder Neues finden.

<http://www.ids-mannheim.de/grammis>